Characterization of 582 natural and synthetic terminators and quantification of their design constraints

Ying-Ja Chen[1], Peng Liu[1], Alec A K Nielsen[1], Jennifer A N Brophy[1], Kevin Clancy[2], Todd Peterson[2] & Christopher A Voigt[1]

Large genetic engineering projects require more cistrons and consequently more strong and reliable transcriptional terminators. We have measured the strengths of a library of terminators, including 227 that are annotated in *Escherichia coli*—90 of which we also tested in the reverse orientation—and 265 synthetic terminators. Within this library we found 39 strong terminators, yielding >50-fold reduction in downstream expression, that have sufficient sequence diversity to reduce homologous recombination when used together in a design. We used these data to determine how the terminator sequence contributes to its strength. The dominant parameters were incorporated into a biophysical model that considers the role of the hairpin in the displacement of the U-tract from the DNA. The availability of many terminators of varying strength, as well as an understanding of the sequence dependence of their properties, will extend their usability in the forward design of synthetic cistrons.

Transcription is the process by which RNA polymerase (RNAP) is recruited to a promoter, mRNA is synthesized and RNAP dissociates at a terminator sequence. Considering the flux of RNAP in this process, we observe that designs emerging from genetic engineering often differ from natural genetics in two ways. First, very strong promoters are frequently required for synthetic circuits. Second, designs are modularly organized along a relatively short stretch of linear DNA. This can create a highly punctate pattern of gene expression, in which strong promoters generate a high flux of RNAP, which then transcribes a relatively short stretch of DNA. This high flux then needs to be sharply stopped so as to not interfere with the next transcription unit. This hard start–hard stop design introduces a need for many strong terminators. This need increases for larger designs: for example, genetic programs that consist of many integrated genetic circuits[1] or multigene pathways that are controlled as individual cistrons[2]. This need is exacerbated by the fact that strong termination requires the connection of multiple terminators in series and because reuse of genetic parts within a design can lead to homologous recombination[3].

Here we characterized a large library of natural and synthetic *E. coli* intrinsic (ρ factor–independent) terminators. Intrinsic terminators consist of a simple motif in the nascent transcript: a short RNA hairpin followed by a U-rich sequence (U-tract), both of which are required for efficient termination[4]. The canonical model of termination is that the hairpin must form within the RNA exit channel of the RNAP[5,6]. Because U:A is the weakest RNA:DNA base pair[7–10], the U-tract causes transcription to pause[11,12], thus allowing the hairpin to form and disrupt the binding between the U-tract RNA and the template DNA[5,11]. Computational models have been developed to search for intrinsic terminators in prokaryotic genomes through identification of hairpins near regions of high U content[13–21]. In addition, biophysical models have been developed to predict how the sequence of a terminator affects its strength[14,22–24].

This study was initiated in response to demand: the larger synthetic constructs (comprising 10–100 promoters) that we were building for temporal control of pathway expression required many more terminators than were available. To address this, we initially constructed a modest library of ~60 natural terminators, but we found that few were comparable in strength to the terminators commonly used in synthetic constructs (such as BBa_B0010)[25]. We therefore built the entire set of 227 annotated terminators from *E. coli*[26]. Although this set covers a wide range of strengths, only a few could be considered very strong. We then made various synthetic libraries that permute the sequence features that contribute to terminator strength. These efforts collectively provide a large data set from which the relative contributions of the different sequence features can be determined.

RESULTS

Characterization of a library of natural terminators

An initial library of 317 natural intrinsic terminators was compiled from various sources including RegulonDB 6.4 (ref. 27), Yager and von Hippel[22] and the Registry of Standard Biological Parts (http://partsregistry.org/) (Online Methods, **Supplementary Table 1** and **Supplementary Note 1**). A subset of 93 terminators was included in the reverse direction (indicated by the suffix '-R'). Terminator strength (T_S) was determined from a simple assay that compares the expression of two fluorescent reporters, one placed before and one after the terminator[28] (**Fig. 1a**). T_S was

[1]Department of Biological Engineering, Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Synthetic Biology Research & Development, Life Technologies, Carlsbad, California, USA. Correspondence should be addressed to C.A.V. (cavoigt@gmail.com).

RECEIVED 24 JANUARY; ACCEPTED 6 MAY; PUBLISHED ONLINE 2 JUNE 2013; DOI:10.1038/NMETH.2515

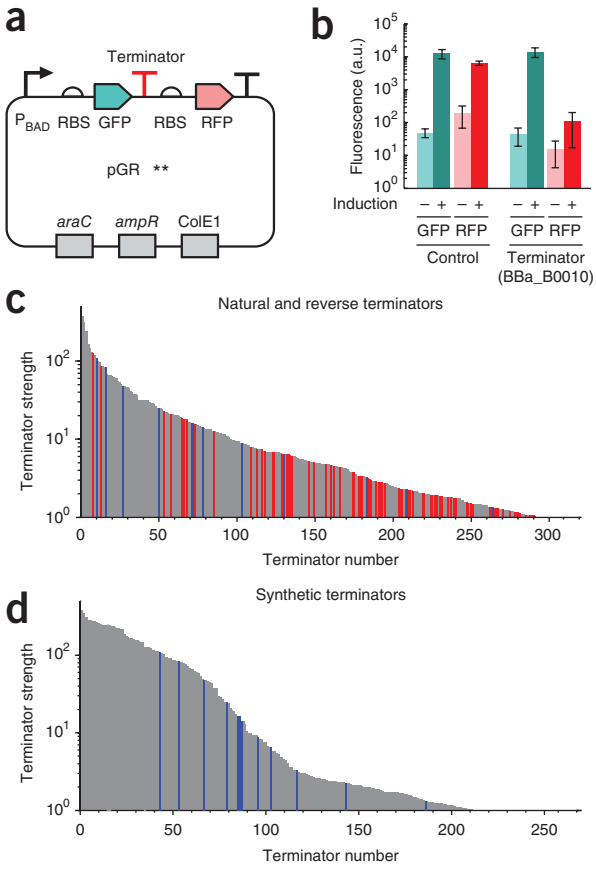

Figure 1 | Measurement of terminator strength for the natural and synthetic libraries. (**a**) Terminators are inserted between two fluorescent proteins in the reporter plasmid. The ** in the plasmid name is a placeholder for the specific terminator (**Supplementary Tables 2–4**). RBS, ribosome-binding site. (**b**) Fluorescence from GFP and RFP are measured in the uninduced (–) and induced (+) states. The control plasmid lacks a terminator. Error bars are s.d. calculated from at least three measurements performed on different days. a.u., arbitrary units. (**c**,**d**) Terminator strengths measured for the natural (**c**) and synthetic (**d**) libraries. The data and sequences for each terminator are provided in **Supplementary Tables 2** and **3**. Commonly used terminators from the Registry of Standard Biological Parts are shown in blue as a reference. Terminators measured in the reverse orientation (-R) are in red.

calculated by applying equation (S1) (**Supplementary Note 1**) and using the fluorescence data, and the plasmid with no terminator insert was used as reference ($T_S = 1.00 \pm 0.36$ s.d.) (**Fig. 1b** and **Supplementary Note 1**). We used this assay to measure the strengths of the complete library of *E. coli* natural and reverse (**Fig. 1c** and **Supplementary Fig. 1a**) and synthetic terminators (**Fig. 1d** and **Supplementary Fig. 1b**; data and sequences are in **Supplementary Tables 2** and **3**, and terminator libraries are available at Addgene (Online Methods)). A small fraction of terminators exhibited unusual expression patterns and were removed (**Supplementary Note 2**, **Supplementary Table 4** and **Supplementary Figs. 2–7**). The strongest forward terminator from the Registry was BBa_B0010 ($T_S = 84 \pm 12$), which forms the basis for many of the strong double terminators (two terminators placed in tandem) used in synthetic constructs. In the library, there are 11 stronger *E. coli* terminators. There are 87 forward terminators with $T_S > 10$, corresponding to terminator efficiencies greater than 90%.

Determination of terminator design constraints

We used the data from the library of *E. coli* forward terminators to identify the sequence features leading to their strengths. The major contributors to terminator strength are discussed here, and additional parameters are discussed in **Supplementary Notes 3–5** (**Supplementary Figs. 8–15**). The strongest contributor is the U-tract, which encompasses the 8 nucleotides immediately downstream of the hairpin[5,11,14] (**Fig. 2a**). Data are shown for the probability of observing a U at each position in the U-tract for those terminators that are strong ($T_S > 40$) or weak ($T_S < 3$) and for the complete library (**Fig. 2b**). There was a notable decline along the U-tract in the probability of observing a U that went from near unity (for strong terminators) to the probability expected from a random distribution of nucleotides at position 11. Among strong terminators there was a sharp drop in the bias toward U that occurred at position 7.

The biochemical role of the U-tract is to provide a weak base-pairing to the template DNA that favors dissociation. The free energy of binding between the U-tract and the template DNA, ΔG_U, can be calculated using

$$\Delta G_U = \Delta G^0_{RNA:DNA} + \sum_{i=1}^{N_U - 1} \Delta G_{RNA:DNA}(n_i, n_{i+1}) \qquad (1)$$

where $N_U = 8$ is the length of the U-tract, $\Delta G^0_{RNA:DNA}$ is the initiation term for RNA:DNA hybridization and $\Delta G_{RNA:DNA}(n_i, n_{i+1})$ is the free-energy contribution of the RNA:DNA hybridization from the dinucleotide pair at positions i and $i + 1$ (ref. 8). As expected, terminator strength correlated with ΔG_U: strong terminators had higher values, thus favoring dissociation (**Fig. 2c**).

Hairpins are a central feature of terminators and serve as the basis for computational methods that search genomes for terminators[13,15,17–19,21]. To determine the free energy of hairpin folding (ΔG_H), we used KineFold to simulate the kinetic folding of mRNA as it is transcribed by RNAP[29] and then Vienna RNA[30] to calculate ΔG_H for the kinetically identified structure. There was a weak correlation between ΔG_H and T_S and no correlation between T_S and loop or stem length (**Supplementary Note 3** and **Supplementary Fig. 8**). However, there was a correlation between T_S and the free energy for the closure of the hairpin loop, ΔG_L (**Fig. 2d**). Strong terminators had a low ΔG_L, meaning that it is easier to form the loop, which suggests a kinetic mechanism that favors rapid loop closure. This is consistent with the observation that the most highly observed terminator loop size is a tetraloop, which increases the stability and rate of folding[31], and with an inverse correlation between T_S and stem mismatches (**Supplementary Note 3**).

We observed a correlation between strength and the base content of the hairpin stem, but it was not equal at every position. Among strong terminators, the GC content was enriched only at the base of the stem (**Fig. 2e**) and not near the hairpin loop (**Fig. 2f**). This is consistent with a previous hypothesis that the free energy released from the base-pairing at the base of the stem contributes the most to the ratcheting of the U-tract off of the DNA[5,32]. The free energy of the base (ΔG_B) is the sum of the base-pair energies of the b1–b3 nucleotides. Stronger terminators had a lower value of ΔG_B (**Fig. 2g**).

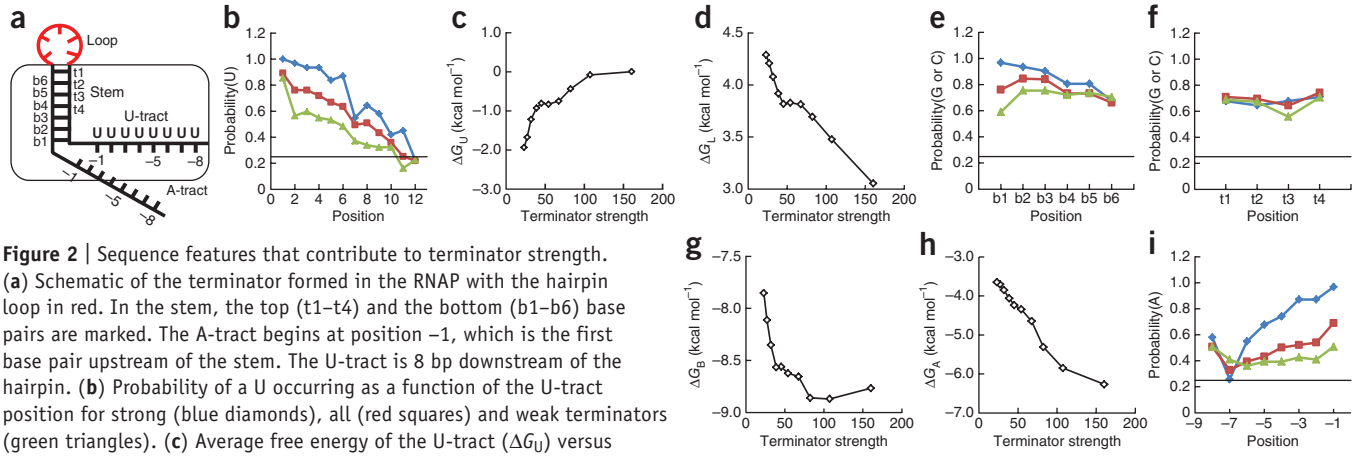

Figure 2 | Sequence features that contribute to terminator strength. (**a**) Schematic of the terminator formed in the RNAP with the hairpin loop in red. In the stem, the top (t1–t4) and the bottom (b1–b6) base pairs are marked. The A-tract begins at position –1, which is the first base pair upstream of the stem. The U-tract is 8 bp downstream of the hairpin. (**b**) Probability of a U occurring as a function of the U-tract position for strong (blue diamonds), all (red squares) and weak terminators (green triangles). (**c**) Average free energy of the U-tract (ΔG_U) versus terminator strength, T_S (representing the average T_S for subsets of the terminator library). (**d**) Average free energy for the closure of the hairpin loop (ΔG_L) plotted against average T_S. (**e**) Probability of either a G or C occurring at the indicated stem position for strong, all and weak terminators. (**f**) Probability of either a G or C occurring at each stem position near the loop for strong, all and weak terminators. (**g**) Average free energy of the base (ΔG_B) versus average T_S. (**h**) Average free energy of the extended hairpin (ΔG_A) as a function of the average T_S. (**i**) Probability of an A occurring at the indicated position for strong, all and weak terminators. For all averages in **c**,**d**,**g**,**h**, the leftmost point represents the complete library and the rightmost point is the strongest 15 terminators. For all probabilities in **b**,**e**,**f**,**i**, strong and weak terminators are defined as $T_S > 40$ and $T_S < 3$, respectively.

There was also a correlation between T_S and the presence of an A-rich tract upstream of the hairpin (**Fig. 2h**). This has often been associated with the bidirectionality of the terminator, that is, the terminator's ability to stop polymerases transcribing from either direction. We determined the probability of observing an A at each position in the A-tract (**Fig. 2i**). There was a sharp increase in the probability of observing an A for strong terminators in the positions close to the base of the hairpin. There was an inversion in the probabilities at the –7 position. Notably, this corresponds to the +7 dip observed for the U-tract, and strong terminators are enriched in complementary nucleotides at these positions.

The role of the A-tract could be to extend the hairpin stem, which contributes to the ratcheting of the U-tract from the DNA. The free energy of the extended hairpin ΔG_A can be calculated as $\Delta G_{HA} - \Delta G_H$, where ΔG_{HA} is the free energy of the folded RNA beginning 8 nucleotides upstream of the hairpin and ending 8 nucleotides downstream of the hairpin. Stronger terminators tended to have more negative ΔG_A, indicating A-tract base-pairing with the U-tract (**Fig. 2h**). Note that ΔG_A is not necessarily minimized by having a poly(A) sequence (unlike with Us and ΔG_U); rather, it is lower when an A-tract is the complement of the U-tract sequence. We characterized a subset of 90 terminators in the opposite orientation but found only a few that exhibited moderate and equivalent termination in both directions (**Supplementary Note 4** and **Supplementary Fig. 14**). It may be that the A-tract does not serve to promote bidirectional termination but rather is involved in forward terminator function for at least a class of very strong terminators.

Design and characterization of synthetic terminators

We designed three libraries of synthetic terminators to dissect the specific contributions of each sequence feature. The first two libraries contained sequences that were synthetic (that is, not derived from a natural source) and did not produce any strong terminators (**Supplementary Note 6** and **Supplementary Figs. 16** and **17**). A third library was then designed by systematically varying each feature but using only those sequences gleaned from strong terminators. We selected three strong natural terminators as scaffolds (**Fig. 3a**). We then swapped into the scaffold features from other strong terminators, including the hairpin loop, the hairpin stem, the A-tract, the U-tract (+4 base pairs (bp)) and both the A- and U-tracts (**Fig. 3b**). Changes in the sequence features resulted in unexpected differences in the robustness of the

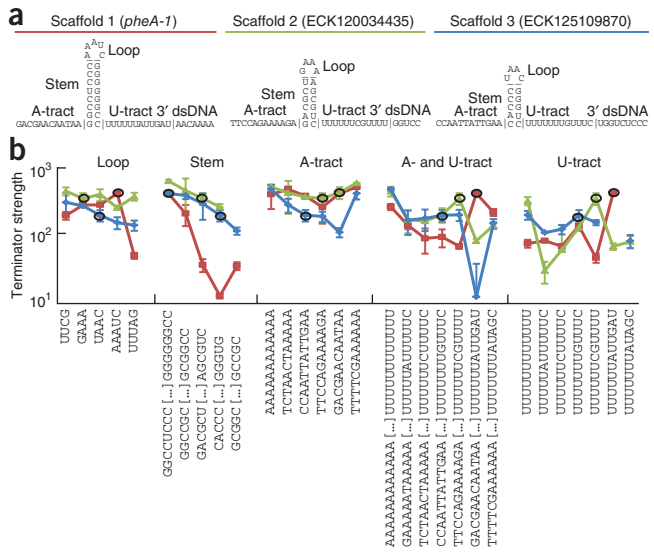

Figure 3 | Impact of changing the various components of a terminator in three strong terminator 'scaffolds'. (**a**) Three scaffolds were used to generate a library of strong terminators. The sequence of each scaffold is divided into the A-tract, stem, loop and U-tract (separated by vertical lines between nucleotides). (**b**) For each scaffold, the component listed at the top of the graph consists of the sequence at the bottom. The colors of the lines correspond to the scaffolds: *pheA-1* (red squares), ECK120034435 (green triangles) and ECK125109870 (blue diamonds). The circled data points represent values for the unperturbed scaffolds. Error bars are s.d. calculated from three measurements performed on different days.

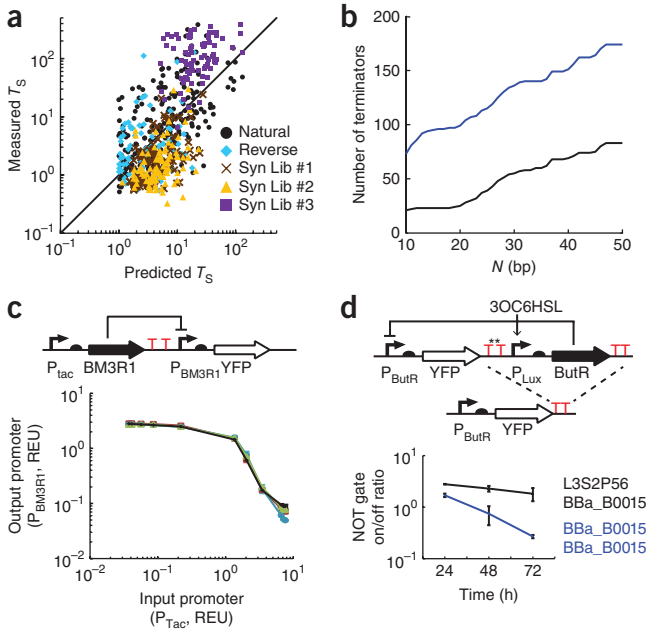

Figure 4 | Biophysical model of terminator strength, T_S, and recombination propensity. (**a**) Measured T_S plotted against the T_S predicted by the model. Terminators with a predicted T_S <1 are omitted from the figure. The fit to the $x = y$ line for all terminators is $R^2 = 0.40$. Syn Lib, synthetic library. (**b**) The largest subset of terminators that share at most N base pairs in sequence identity plotted against N for strong (T_S >50, black) and moderate-to-strong (T_S >10, blue) terminators. (**c**) Transfer function for the NOT gate measured when one of three terminators (L3S3P22, red square; ECK120029600, blue circle; L3S3P00, green triangle) was substituted as compared to the original terminators (BBa_B0015, black diamond). Error bars are the s.d. of three biological replicates measured on the same day. (**d**) Schematic and results of a NOT gate tested for recombination. The terminators being replaced are labeled with **. When the BBa_B0015 terminator was used repeatedly, nearly all of the plasmid showed recombination, and the function of the NOT gate was compromised as indicated by the on/off ratio. When the L3S2P56 terminator was substituted into the terminator position in red, recombination was minimal, and the on/off ratio of the NOT gate was maintained. REU, relative expression units.

scaffolds. For most features, the scaffolds reacted differently to changes: in other words, there generally was no optimal sequence for each feature. Notably, some substitutions were disruptive to strength despite that all of the sequences originated from strong terminators.

Each scaffold reacted differently to changes in the U-tract. Substituting a perfect U-tract (that is, a U-tract containing only Us) improved the strength of scaffold 2 and 3 but reduced the strength of scaffold 1 by sixfold. With the exception of the perfect U-tract, the optimal sequence was that associated with the wild-type terminator; substitutions from this sequence tended to be detrimental. The A-tract in general had the least effect, but one A-tract containing five As improved all three terminators. The perfect A-tract improved scaffolds 2 and 3. Scaffold 3, which had the smallest hairpin stem, was the most affected by A-tract substitution and had the largest improvement due to better A-tracts. Changing the A-tract simultaneously with the U-tract almost always recovered some of the loss in strength that occurred with the substitution of the U-tract alone. The exception to this was the double A- and U-tract substitution for scaffold 1, which lowered the terminator activity of the other scaffolds.

The impact of the hairpin on scaffolds 2 and 3 was as expected. On average, the best loop was GAAA, which forms a stable hairpin[31]. All of the terminators were strongest with the longest 8-bp stem region and were weaker with shorter stems. Scaffold 1 reacted very differently to substitutions across all features, but this sensitivity was especially apparent with the hairpin, when the stem or loop diverged from their native sequences. It was also the only scaffold for which substituting in a perfect U-tract lowered terminator strength. The secondary structure of the hairpin is predicted to be correct in the cases of both the original and the perfect U-tract, but if the native U-tract is present, the hairpin is not as extended with the perfect A-tract as it is with the native A-tract. We overcame this problem by substituting in a perfect A-tract, which leads to the formation of a long, stable hairpin, and the strength of the terminator was recovered.

Scaffold 1 ('*pheA-1*') was one of the strongest in the library ($T_S = 244 \pm 46$), but it was also the most sensitive to changes in its sequence. The hairpin is predicted to form a pseudoknot with the region immediately after the U-tract (**Supplementary Note 7** and **Supplementary Fig. 18**). The terminator is very sensitive to short hairpin stems, and if the loop or the U-tract are changed in ways that disrupt the pseudoknot, then strength is lost. This terminator is unique in our library because it occurs upstream of the *pheA* gene and is involved in attenuation[33], a process in which pseudoknots have been cited to be important for regulatory function[34]. In addition, pseudoknots are known to cause ribosomal 'roadblocks', thus decreasing translational coupling[35]. The model predicts an eightfold lower strength for this terminator than what we measured, so this higher T_S may represent the pseudoknot's combined impact on the transcriptional and translational levels.

Biophysical model of terminator function

We developed a simple biophysical model to capture how the sequence of a terminator affects its strength. This requires a formalism to connect the contributions from the features that are observed to be important, including the U-tract ΔG_U, the hairpin loop ΔG_L, the stem base ΔG_B and the A-tract ΔG_A. The model is based on the hybrid-shearing mechanism (**Supplementary Note 8**). It considers the termination efficiency as a probability that RNAP will dissociate upon reaching the terminator. This is assumed to be a two-step process in which the hairpin has to nucleate and then the U-tract is ratcheted from the RNAP. Termination occurs when these steps occur faster than the RNAP progressing through the terminator[4,23]. Considering this mechanism, we derived a kinetic model (**Supplementary Fig. 19**) to calculate the terminator strength, yielding

$$T_S = 1 + \frac{1}{B_1 e^{\beta_1 \Delta G_L} + B_4 e^{\beta_4 (\Delta G_B + \Delta G_A - \Delta G_U)} (1 + B_1 e^{\beta_1 \Delta G_L})} \quad (2)$$

where $B_1 = 0.005$, $B_4 = 6.0$, $\beta_1 = 0.6$ and $\beta_4 = 0.45$ are fit parameters (**Supplementary Fig. 20**). The predicted T_S is calculated by equation (2) with the free energies provided by equations (1), (S7) and (S8) (**Supplementary Note 8**). We used equation (2) to fit the experimentally derived terminator strengths for all 582 terminators (**Fig. 4a**).

Terminator function in the context of a genetic circuit

The availability of a library of many strong terminators enables larger genetic systems to be constructed that are composed of multiple cistrons. Typically, double terminators are used to stop transcription. Because these parts can be up to ~168 bp of DNA, this use of double terminators can lead to homologous recombination when used at multiple locations in a design. This poses a challenge for any multidevice system and is particularly problematic in building programs that involve connecting transcriptional genetic circuits, each of which consists of multiple terminators[1,36]. In *E. coli*, homologous recombination occurs most frequently when there is a contiguous stretch of sequence >25 bp (refs. 37–39). Considering the complete library of natural and synthetic terminators, we calculated that the largest subset is 39 terminators that are both strong (T_S >50) and do not share a contiguous sequence larger than 25 bp (**Supplementary Note 9** and **Supplementary Fig. 21**). For moderate-to-strong terminators (T_S >10), this subset increased to 115 (**Fig. 4b**).

We then tested how substituting the terminators affected the function of a genetic circuit. A transcriptional NOT gate was chosen, as these have frequently appeared in genetic programs. NOT gates consist of an input promoter that drives the expression of a repressor that turns off an output promoter. Two strong terminators are traditionally required between the repressor and the output promoter. The repeated use of terminators for both the repressor and the output reporter has previously resulted in recombination[3]. First, we replaced the BBa_B0015 (BBa_B0010-BBa_B0012 pair) terminators with three different single terminators from our library. These substitutions had a negligible impact on the input-output relationship of the circuit (**Fig. 4c**). Finally, we had noticed a problem with recombination for a genetic circuit that we had been building due to a repressor (ButR) that exhibited some toxicity. We substituted one of the BBa_B0015 terminators with the L3S2P56 terminator from our library and performed experiments[3] to demonstrate that, over 70 generations, the gate was stable and recombined variants were undetectable (**Fig. 4d**, **Supplementary Note 10** and **Supplementary Fig. 22**).

DISCUSSION

The precise control of RNAP flux is a challenge in genetic engineering. Designs frequently have multiple promoters in various orientations. The ability to reliably stop transcription at a defined point is essential for ensuring that different regions of a design do not interfere with each other. To date, only a few strong terminators have been available, and these have been widely reused within designs and across projects. There are also many cases in which weak terminators have been inadvertently used, which has unintended consequences. For example, a weak terminator that is in common use is from phage and has an efficiency of only 50–70% (T_S = 2–3)[12,40], meaning that half of the polymerase flux continues beyond the terminator. Our large library of natural and synthetic terminators enabled us to extract the contributions of the features based on averages; we were thus not limited to the particular mechanism or idiosyncrasies of a single terminator. The next step toward fully using terminators as a more prominent design component of synthetic biology projects will be to better understand the impact of the surrounding genetic context and how the 3′ region of cistrons, including terminators, can be used as a tuning knob to obtain higher-precision expression control.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

C.A.V., P.L., A.A.K.N., J.A.N.B. and Y.-J.C. are supported by Life Technologies and the US National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). Y.-J.C. thanks the PhRMA Foundation for a Postdoctoral Fellowship in Informatics. We thank R. Landick and D.Y. Zhang for their advanced review for the manuscript.

AUTHOR CONTRIBUTIONS

C.A.V. conceived and supervised the project. Y.-J.C., A.A.K.N. and J.A.N.B. designed and performed the experiments. C.A.V., P.L. and Y.-J.C. constructed the biophysical model. K.C. and T.P. oversaw the project. C.A.V., Y.-J.C. and P.L. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C. & Voigt, C.A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
2. Watanabe, K. *et al.* Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat. Chem. Biol.* **2**, 423–428 (2006).
3. Sleight, S.C., Bartley, B.A., Lieviant, J.A. & Sauro, H.M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4**, 12 (2010).
4. Peters, J.M., Vangeloff, A.D. & Landick, R. Bacterial transcription terminators: the RNA 3′-end chronicles. *J. Mol. Biol.* **412**, 793–813 (2011).
5. Komissarova, N., Becker, J., Solter, S., Kireeva, M. & Kashlev, M. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol. Cell* **10**, 1151–1162 (2002).
6. Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S. & Nudler, E. An allosteric path to transcription termination. *Mol. Cell* **28**, 991–1001 (2007).
7. Martin, F.H. & Tinoco, I. DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.* **8**, 2295–2299 (1980).
8. Sugimoto, N. *et al.* Thermodynamic parameters to predict stability. *Biochemistry* **34**, 11211–11216 (1995).
9. Huang, Y., Chen, C. & Russu, I.M. Dynamics and stability of individual base pairs in two homologous RNA-DNA hybrids. *Biochemistry* **48**, 3988–3997 (2009).
10. Huang, Y., Weng, X. & Russu, I.M. Structural energetics of the adenine tract from an intrinsic transcription terminator. *J. Mol. Biol.* **397**, 677–688 (2010).
11. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).
12. Macdonald, L.E., Zhou, Y. & McAllister, W.T. Termination and slippage by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **232**, 1030–1047 (1993).
13. Argaman, L. *et al.* Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**, 941–950 (2001).
14. d'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* **216**, 835–858 (1990).
15. de Hoon, M.J.L., Makita, Y., Nakai, K. & Miyano, S. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput. Biol.* **1**, e25 (2005).
16. Gardner, P.P., Barquist, L., Bateman, A., Nawrocki, E.P. & Weinberg, Z. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845–5852 (2011).

17. Kingsford, C.L., Ayanbule, K. & Salzberg, S.L. Rapid, accurate, computational discovery of rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* **8**, R22 (2007).
18. Lesnik, E.A. *et al.* Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* **29**, 3583–3594 (2001).
19. Mitra, A., Angamuthu, K., Jayashree, H.V. & Nagaraja, V. Occurrence, divergence and evolution of intrinsic terminators across eubacteria. *Genomics* **94**, 110–116 (2009).
20. Mitra, A., Kesarwani, A.K., Pal, D. & Nagaraja, V. WebGeSTer DB—a transcription terminator database. *Nucleic Acids Res.* **39**, D129–D135 (2011).
21. Unniraman, S., Prakash, R. & Nagaraja, V. Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.* **30**, 675–684 (2002).
22. Yager, T.D. & von Hippel, P.H. A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry* **30**, 1097–1118 (1991).
23. von Hippel, P.H. & Yager, T.D. Transcript elongation and termination are competitive kinetic processes. *Proc. Natl. Acad. Sci. USA* **88**, 2307–2311 (1991).
24. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41**, 5139–5148 (2013).
25. Kwon, Y.S. & Kang, C. Bipartite modular structure of intrinsic, RNA hairpin-independent termination signal for phage RNA polymerases. *J. Biol. Chem.* **274**, 29149–29155 (1999).
26. Gama-Castro, S. *et al.* RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39**, D98–D105 (2011).
27. Gama-Castro, S. *et al.* RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* **36**, D120–D124 (2008).
28. Huang, H. *Design and Characterization of Artificial Transcriptional Terminators*. Master's thesis, Massachusetts Institute of Technology (2007).
29. Xayaphoummine, A., Bucher, T. & Isambert, H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* **33**, W605–W610 (2005).
30. Hofacker, I.L. *et al.* Fast folding and comparison of RNA secondary structures. *Monatshefte Chemie* **125**, 167–188 (1994).
31. Varani, G. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379–404 (1995).
32. Larson, M.H., Greenleaf, W.J., Landick, R. & Block, S.M. Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**, 971–982 (2008).
33. Zurawski, G., Brown, K., Killingly, D. & Yanofsky, C. Nucleotide sequence of the leader region of the phenylalanine operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **75**, 4271–4275 (1978).
34. Lopatovskaya, K.V., Seliverstov, A.V. & Lyubetsky, V.A. Attenuation regulation of the amino acid and aminoacyl-tRNA biosynthesis operons in bacteria: a comparative genomic analysis. *Mol. Biol.* **44**, 128–139 (2010).
35. Tholstrup, J., Oddershede, L.B. & Sørensen, M.A. mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res.* **40**, 303–313 (2012).
36. Siuti, P., Yazbek, J. & Lu, T.K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).
37. Fujitani, Y., Yamamoto, K. & Kobayashi, I. Dependence of frequency of homologous recombination on the homology length. *Genetics* **140**, 797–809 (1995).
38. Shen, P. & Huang, H.V. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**, 441–457 (1986).
39. Lovett, S.T., Luisi-DeLuca, C. & Kolodner, R.D. The genetic dependence of recombination in *recD* mutants of *Escherichia coli*. *Genetics* **120**, 37–45 (1988).
40. Temme, K., Hill, R., Segall-Shapiro, T.H., Moser, F. & Voigt, C.A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* **40**, 8773–8781 (2012).

ONLINE METHODS

Plasmid construction. The terminator of interest was inserted between the genes encoding green fluorescent protein (GFP) and red fluorescent protein (RFP); expression was driven by a single P_{BAD} promoter inducible by arabinose. We cloned elements into a plasmid with ColE1 origin and ampicillin resistance by removing the RNase E sites from pSB1A10. Plasmid backbone was prepared by restriction digestion with EcoRI and SpeI. For 50 µg, 100 units of EcoRI-HF (New England Biolabs R3101) and SpeI (New England Biolabs R0133) were combined in 1× NEBuffer 4 (New England Biolabs B7004) and 1× BSA (New England Biolabs B9001), digested at 37 °C for 1 h and inactivated at 80 °C for 20 min. The digested plasmids were extracted after gel electrophoresis using the QIAquick Gel Extraction Kit (Qiagen) following standard protocols.

Terminator inserts were ordered as oligos for both forward and reverse strands with the compatible restriction digestion sites built into the oligos (AATTC on the 5′ end and A on the 3′ end for the forward strand; CTAGT on the 5′ end and G on the 3′ end of the reverse strand). The forward and reverse oligos were phosphorylated in a solution containing 400 nM of each of the oligos, 1× T4 DNA ligase buffer (Life Technologies 46300-018) and 100 units of T4 polynucleotide kinase (New England Biolabs M0201) combined in 100 µl and then incubated at 37 °C for 30 min. Immediately following phosphorylation, equal volume (100 µl) of 6× SSC (150 mM sodium chloride, 15 mM sodium citrate) was added to the solution to anneal the oligos by denaturing the oligos at 95 °C for 5 min and cooling the temperature at 0.1 °C/s until the temperature reached 12 °C.

Terminator inserts were ligated to the digested plasmid backbone by combining 32 ng (20 fmol, 4 nM) of digested plasmid backbone with 1 µl of 200 nM annealed and phosphorylated insert in 5 µl solution containing 1× T4 DNA ligase buffer and 1 U T4 DNA Ligase (Life Technologies 15224-017) at room temperature for 30 min. The ligation products were immediately transformed into 50 µl Z-competent (ZymoResearch T3002) *E. coli* DH5α cells thawed on ice. The competent cells were prepared from cells grown to mid-log phase (OD_{600} = 0.55) at 18–20 °C following standard protocols provided by the vendor and stored at −80 °C. The transformed cells were diluted by 5–200× in SOC medium (SOB, 20 mM glucose) to yield approximately five colonies when 5 µl of solution was spotted onto an agar plate. Single colonies were picked to perform colony PCR using the GoTaq polymerase and master mix (Promega) following standard protocols and using the following primers: qPCR_GFP_2_FW (TGGAAGCGTTCAACTAGCAG) and qPCR_RFP_2_RV (AGTCTTCGTTGTGGGAGGTG), which yields a product of ~800 bp containing the insert region. The products were sequenced using the primer qPCR_insert_1_FW (GTCCACACAATCTGCCCTTT). Cells containing the correct clones from the sequencing results were selected and used for the flow cytometry assay.

The plasmids (**Supplementary Note 11** and **Supplementary Fig. 23**) are available upon request at Addgene (http://www.addgene.org/Syn_Bio/Voigt/Terminators/).

Cell growth and induction. All cloning and assays for terminator strength were performed in *E. coli* DH5α strains. Cells were inoculated in 200 µl Luria-Bertani (LB)-Miller medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, Fisher Scientific) with 100 µg/ml ampicillin in a 96-well plate covered with a breathable membrane (AeraSeal, Excel Scientific) at 37 °C, 1,000 r.p.m. for 16 h in a Digital Thermostatic Shaker DTS-4 (Elmi). Overnight culture was diluted 200-fold by mixing 1 µl culture into 199 µl of LB medium containing 10 mM L-arabinose and 100 µg/ml ampicillin. After 3 h of induction, 15 µl of culture was added to 185 µl of 1× PBS with 2 mg/ml kanamycin for flow cytometry measurement.

Flow cytometry. Measurements were taken using the LSRFortessa flow cytometer (BD Biosciences). The voltage gains for each detector were set so that the full dynamic range was used for a control specimen expressing *GFP* and *RFP* without any terminators in between: FSC, 700 V; SSC, 241 V; FITC, 407 V; PE-TxRed, 650 V. Compensation was set by measuring cells that express only *GFP* or *RFP*. There was no spectral overlap detected from FITC to PE-TxRed. The spectral overlap from PE-TxRed to FITC was 0.11%. Thirty thousand events gated by FSC-H and FSC-W to contain most nonaggregating live cells were collected at 0.5 µl/s sample flow rate under high-throughput mode. Data from FITC and PE-TxRed channels were extracted as the GFP and RFP data. Flow cytometry data were analyzed using FlowJo 7.6.5 (Tree Star). FSC-A and SSC-A were used to gate live cells containing 50–70% total cells. The geometric means in the FITC and PE-TxRed channels were exported as the fluorescence in GFP and RFP.

Transfer-function measurement. We substituted different terminators into a NOT-gate plasmid containing a P_{tac} promoter driving the BM3R1 repressor, which represses the output of YFP. In this construct, a BBa_B0015 double terminator is used to terminate transcription from the BM3R1 repressor so that it does not read through into YFP. The ampicillin resistance gene resides downstream of YFP, followed by the p15A plasmid origin. Constitutive promoters drive LuxR and LacI in the opposite direction. NOT-gate plasmids with different terminators were created by first adding restriction sites to the original BM3R1 NOT-gate plasmid by PCR with the following primers: BM3R1_ES_fw (ACAGAACTAGTTCTGCGTTTATACGACGTACGGTGGAA) and BM3R1_ES_rv (ACTGAGAATTCGCCCTAGTTAGCTCTG ACGGCTCA). The PCR product was digested with DpnI (New England Biolabs R0176S) for 1 h at 37 °C to remove the plasmid template and then purified using a Zymo PCR purification kit (Zymo Research D4003). Purified PCR product was then digested with EcoRI and SpeI as described above and annealed, and phosphorylated oligonucleotides containing the terminators were then ligated into the digested PCR backbone. The resultant plasmids were then transformed into *E. coli* DH10B and screened using colony PCR and sequence verified. Cells containing the correct clones from the sequencing results were selected and used for the flow cytometry assay.

E. coli containing NOT-gate plasmids were grown as described previously[41]. First, cells were inoculated from single colonies on LB agar plates and grown overnight in 4 ml LB in Falcon tubes at 37 °C with shaking (250 r.p.m.). Then the overnight cultures were diluted 200-fold into prewarmed M9 medium (Sigma-Aldrich) supplemented with 2 mM $MgSO_4$ (Fisher Scientific), 100 µM $CaCl_2$ (Sigma), 0.4% glucose (Sigma), 0.2% casamino acids (Acros), 34 ml × 10 mg/ml (for 1 liter) thiamine (vitamin B1). The 96-well plates were then incubated at 37 °C in a digital thermostatic shaker with 1,000 r.p.m. shaking for 3 h until the OD_{600}

was ~0.12–0.14. The cultures were then diluted 700-fold and loaded into prewarmed M9 medium containing inducer in 96-well plates. These new plates were incubated at 37 °C in a digital thermostatic shaker with 1,000 r.p.m. shaking for 6 h to maintain exponential growth. Finally, after the 6-h induction, a sample of each culture was transferred to a new plate containing 150 ml PBS and 2 mg/ml kanamycin to stop protein expression.

Fluorescence was measured by flow cytometry. The geometric mean of each sample was calculated. The autofluorescence of *E. coli* DH10B cells containing no plasmid was measured and subtracted from the mean value to generate the fluorescence values reported. Each fluorescence value was then divided by the fluorescence of *E. coli* DH10B cells containing a YFP reference plasmid (J23101_BBa-B0034-YFP) to obtain relative expression units (REU).

Recombination experiments. We studied recombination in a genetic NOT gate comprising two cistrons: a ButR-repressible promoter (P_{ButR}) driving expression of YFP followed by an HSL-inducible promoter (P_{Lux}) driving expression of the slightly toxic repressor ButR. We designed a repeated-terminator version in which both cistrons are terminated by the BBa_B0015 double terminator, and we also re-engineered versions in which one of the terminators was replaced with a synthetic terminator (L3S2P21 or L3S2P56; see **Supplementary Note 10**). *E. coli* DH5α cells transformed with the repeated-terminator and re-engineered genetic NOT gates were inoculated and grown continuously in liquid LB medium (Teknova) with 50 μg/ml kanamycin sulfate (Life Technologies) in a 37 °C incubator shaking at 250 r.p.m. This growth protocol was carried out in the presence and absence of 20 μM 3OC6HSL inducer to ascertain each NOT gate's off state and on state, respectively. Every 24 h, the cultures were diluted 1:1,000 into fresh LB medium with or without inducer, resulting in approximately ten generations per day ($\log_2(1{,}000) = 9.97$). In addition, every 24 h a sample of cells was diluted 1:200 into 1× PBS and measured via cytometry. Each 72-h evolution experiment was run in triplicate for both the repeated-terminator and different-terminator NOT gates.

In addition, uninduced cultures were grown in the manner described above in the absence of 3OC6HSL for 7 d. Plasmid DNA was harvested from the cultures using a plasmid miniprep kit (Qiagen), and 500 ng of plasmid DNA was restriction digested in 50-μl reactions using 1× NEBuffer 4, 1 μl BamHI, 1 μl AvrII, 1 μl SacII and 0.5 μl 100× BSA (NEB) for 2 h at 37 °C. Digested DNA was purified and concentrated to 25 μl using a DNA purification kit (Qiagen) and then visualized on a 1% agarose gel with 1× SYBR Safe (Invitrogen).

41. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C.A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30**, 1137–1142 (2012).

doi:10.1038/nmeth.2515